Osteoarthritis and Cartilage (2008) **16,** 742–748

doi:10.1016/j.joca.2008.02.021

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI–OMERACT task force

Extended report

L. Gossec M.D.†‡¶¶*, J. M. Jordan M.D., Ph.D.§, S. A. Mazzuca M.D., Ph.D.‖, M.-A. Lam M.D.¶#, M. E. Suarez-Almazor M.D., Ph.D.††, J. B. Renner M.D., Ph.D.§, M. A. Lopez-Olivo M.D.††, G. Hawker M.D., Ph.D.¶#, M. Dougados M.D.†‡ and J. F. Maillefert M.D., Ph.D.‡‡§§‖‖

For the OARSI–OMERACT task force "total articular replacement as outcome measure in OA"

† *Paris Descartes University, Medicine Faculty, UPRES-EA 4058, Paris, France*

‡ *APHP, Rheumatology B Department, Cochin Hospital, Paris, France*

§ *Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC, USA*

‖ *Indiana University School of Medicine, Indianapolis, IA, USA*

¶ *Division of Rheumatology, Department of Medicine, Women's College Hospital, Canada*

Clinical Epidemiology and Health Care Research Program, Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Canada

†† *University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA*

‡‡ *Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France*

§§ *University of Burgondy, Dijon F-21079, France*

‖‖ *INSERM U887, Dijon F-21079, France*

¶¶ *Université Montpellier I, EA 2415 Epidémiologie et Biostatistique, Montpellier, France*

Summary

Objective: The objective of this work was to compare the measurement properties of three categorical X-ray scoring methods of knee osteoarthritis (OA), both on semiflexed and extended views.

Methods: In data obtained from trials and cohorts, X-rays were graded using Kellgren and Lawrence (KL), the OA Research Society International (OARSI) joint space narrowing score, and measurement of joint space width (JSW). JSW was analyzed as a categorical variable. Construct validity was assessed through logistic regression between X-ray stages and Western Ontario and McMaster Universities OA Index. Inter-observer reliability was assessed in 50 subjects for extended views by weighted kappa. Intra-observer reliability and sensitivity to change were assessed separately for extended and semiflexed views in 50 patients who had both views performed, over a 30-month interval, by weighted kappa and standardized response mean (SRM).

Results: Extended views were available from three trials and two cohorts (1759 X-rays), including one trial in which both extended and semiflexed views (antero-posterior) were obtained. Correlation with clinical parameters was low for the three scoring methods, except for the single community-based cohort. Inter-rater reliability was higher for categorical JSW in extended views (kappa, 0.86 vs 0.56 and 0.48 for KL and OARSI, respectively). Intra-rater reliability was higher for categorical JSW, both in extended views (0.83 vs 0.61 and 0.71) and in semiflexed views (0.89 vs 0.50 and 0.67). Sensitivity to change was also higher for categorical JSW, particularly in semiflexed views (SRM, 0.49 vs 0.22 and 0.34).

Conclusion: These results indicate categorical JSW, in particular on semiflexed views, may be the preferred method to evaluate structural severity in knee OA clinical trials.

Key words: Knee, Osteoarthritis, Progression, Radiography, Validity, Reproducibility.

Introduction

Osteoarthritis (OA) is a major cause of disability worldwide[1]. Over the past years, interest has grown among the scientific community, pharmaceutical companies, and regulatory agencies in the development of drugs that might influence the natural history of OA by preventing, retarding, or reversing cartilage breakdown. Interest exists, therefore, in identifying a valid, dichotomous outcome variable that reflects the

*Address correspondence and reprint requests to: Dr Laure Gossec, M.D., Service de Rhumatologie B, Hôpital Cochin, 27 rue du faubourg Saint-Jacques, 75014 Paris, France. Tel: 33-158412606; Fax: 33-143549256; E-mail: laure.gossec@cch.aphp.fr

Received 30 January 2008; revision accepted 14 February 2008.

natural history of OA. In particular, interest has grown in using the requirement of total joint replacement as a "hard" outcome measure[2,3]. Limitations exist, however, in the use of such an outcome, in particular variability in the decision to perform surgery, the length of surgical waiting lists, and responsiveness. Thus, a better alternative might be to change the criteria "time to total joint replacement" to "time to fulfill the criteria for total joint replacement"[4]. In this context and as described elsewhere[5], an international working group was created under the auspices of OA Research Society International (OARSI) and Outcome Measures in Rheumatology Clinical Trials (OMERACT) in order to elaborate a set of criteria defining theoretical requirement for total joint replacement in knee and hip OA, for use in clinical trials evaluating potential disease-modifying drugs and other interventions in OA. It was decided that the domains of pain, physical function and joint structure on X-rays[5] would be combined as a surrogate measure of outcome. For each of these domains, a categorical outcome will be used to render combination of the domains feasible. As a first step, three working subgroups were constituted, to determine which instrument should be used to evaluate these domains.

This article presents the work of the "structure group" on knee OA. The objective was to examine categorical measures of OA defining structural severity in knee OA in terms of suitability to enter into a composite criterion representing a dichotomous indicator for joint replacement. At this time, evaluation of structural degradation in trials uses a quantitative measurement of joint space width (JSW), either at the narrowest point or mean JSW, by precise measurement using a ruler or caliper, or through computer assisted techniques[6]. However, JSW measurement provides a continuous variable, while the working group aimed to establish a dichotomized outcome (virtual indication for joint replacement yes–no). This binary outcome could then be used as a definition for "responder/non-responder" in OA clinical trials (whereas a continuous outcome does not define a responder). To this end, it was necessary to categorize or dichotomize the continuous variable JSW, or change in JSW, or to evaluate the domain structure using a widely used categorical instrument, such as the Kellgren and Lawrence (KL) scoring method[7], or the OA Research Society International (OARSI) joint space narrowing severity grades[8].

At the present time, there is also no consensus on the optimal radiographic technique to evaluate structural severity in knee OA[9]. Researchers generally agree that weight-bearing views are best since supine views may not demonstrate joint space narrowing evident on weight-bearing views[10]. However, different weight-bearing views are used: either extended views, or semiflexed. Although there are some data in favor of semiflexed views[11–14], extended views are still widely used. Among semiflexed views, there are also various techniques available: (1) antero-posterior (fluoroscopically assisted) semiflexed views[15], (2) metatarsophalangeal semiflexed views, which are postero-anterior views, non-fluoroscopically assisted[13], (3) Lyon-Schuss views, which are postero-anterior views, fluoroscopically assisted[16], (4) fixed flexion views with Synaflex positioning device; these techniques all present both advantages and disadvantages[17]. Finally, femoro-patellar views also present unresolved issues, but are not the objective of this study.

As part of the process of this OARSI–OMERACT initiative, it was necessary to assess different radiographic views and categorical measures of structural severity in knee OA. To assess a potential outcome measure, it is necessary to assess its psychometric properties, as defined by the OMERACT filter. The OMERACT filter[18] checks that a potential outcome measure is truthful, i.e., reflects what it is supposed to reflect, and is discriminant, which includes reproducibility, and sensitivity to change, over time, and between different severity stages. The last element in the OMERACT filter refers to feasibility, which is not assessed through statistics. There are numerous published data regarding the psychometric properties of the different views and scoring techniques in knee OA[6,11,13,14,17]. However, to our knowledge, studies in which all the properties of the OMERACT filter are compared head-to-head are lacking, making it difficult to choose the most effective scoring system for our purposes.

The objective of this work was to provide such a head-to-head comparison of the different radiographic views and the different scoring techniques in knee OA, by assessing their psychometric properties[18].

Patients and methods

DATA SOURCES

A call for data (available databases with X-rays either from trials or cohorts) was sent out to the OARSI–OMERACT group and to leading experts. The databases could be trials or cohorts, and the X-rays could have been obtained using extended or semiflexed views. Ideally, the X-rays had been analyzed with all three scoring methods; in all other cases, the X-rays had to be available for interpretation. It was also necessary to obtain some clinical data (see below). Finally, five databases were proposed, they all had the necessary available data (X-rays and clinical data) and all available X-rays were analyzed for construct validity. In one database, the X-rays had been analyzed with the three scoring methods; in the other databases, X-rays were reanalyzed centrally. For intra- and inter-reproducibility and sensitivity to change, a random sample of the available X-rays was analyzed.

Ethics approval was obtained where necessary for reinterpretation of the X-rays.

SCORING OF X-RAYS

For the purposes of this study, at least one radiographic view was analyzed for each knee, using a standing extended view. Where available, a semiflexed view was also analyzed. These views allowed the analysis of the tibio-femoral joint. The patellofemoral joint was not assessed. All available radiographs were analyzed; the radiographs were required to be of sufficient quality to allow interpretation of the joint region. If the magnification was reported, JSW could be measured, otherwise only OARSI and KL scores were assessed. In trials, only one knee per individual was analyzed, the "index knee" (symptomatic or most symptomatic knee). In cohorts, both knees and all four compartments (medial and lateral of both knees) were analyzed and the most severe compartment radiographically was used for analysis.

X-rays were scored three times each: (1) radiologic grade according to the KL classification[7] based on Kellgren's original written description; (2) OARSI grade for joint space narrowing[8], and (3) JSW as a continuous variable. Joint space was measured manually at the narrowest point, with a magnifying lens fitted with a graticule. As no categorization was published before for JSW at the knee, JSW was changed into a categorical variable using the median and quartile values of the first database analyzed: cut-offs were 2 mm (quartile), 3.5 mm (median), and 5 mm (quartile).

Reporting of the films was undertaken by rheumatologists who had all undergone training by other rheumatologists to standardize the X-ray scoring. Most radiographs were read by JFM (Paris, Indianapolis, and Houston). In addition, a set of 50 radiographs were read by LG in order to evaluate inter-reader reliability. LG previously underwent training with JFM. The readers were blinded to all clinical and questionnaire data.

CLINICAL DATA COLLECTION

Demographic data comprising age, sex, race/ethnicity, and body mass index (BMI) were collected at the same timepoint as X-rays were performed. Clinical severity was estimated through the subscales of the Western Ontario and McMaster Universities OA Index (WOMAC)[19]. Each item has five response options (none/mild/moderate/severe/extreme) and yields total subscale scores for pain (five items, total score: 0–20), disability (17 items, total score: 0–68) and stiffness (two items, total score: 0–8). Results were normalized to a 0–100 score.

DATA ANALYSIS

Validity was assessed through the cross-sectional relationship between X-ray stages and WOMAC total, pain, function and stiffness scores, by logistic regression. In databases obtained from trials, the baseline data were used for analysis. This assessment of validity will be termed here "construct validity". Inter-observer reliability was assessed separately for extended and semiflexed view X-rays in 50 subjects for whom both views had been performed by two readers (JFM and LG). Intra-observer reliability was assessed separately for extended and semiflexed view X-rays in the same patients, with one reader at a 48 h interval. Thus the same study participants were used for evaluating intra-observer reproducibility for extended views and semiflexed views.

Sensitivity to change was assessed on 50 pairs of extended view X-rays and 50 pairs of semiflexed view X-rays, in the same study participants (but different participants from reliability) over a 30-month interval. The films were read with knowledge of the order.

STATISTICAL ANALYSIS

Statistical analysis was performed using SAS statistical software (SAS institute Inc, Cary, NC). For construct validity, logistic regression analyses were carried out to model symptomatic severity (WOMAC total, pain, function and stiffness), categorized into quartiles, by X-rays' grades, adjusted for age, race/ethnicity, sex, and BMI (proportional odds model as assumptions were satisfied). Each of the higher quartiles of X-ray grading was compared to the lowest grade. Associations between symptomatic severity on each item and radiographic severity for KL, OARSI joint space narrowing and JSW (categorized) were expressed as adjusted odds ratios with 95% confidence intervals (CIs). The least severe radiographic category was used as the reference category. For reliability, weighted kappas were calculated, as well as intra-class correlation coefficients (ICCs)[20]. Sensitivity to change was assessed by standardized response mean, SRM: mean (month 30−month 0) X-ray score change/standard deviation (SD) of X-ray score change. Although SRM was not developed as a measure for semi-quantitative data, it was used here since the assumption of calculations of mean and SD regarding equal intervals was violated to the same extent by each of the outcomes. Statistical significance was set at 0.05.

Results

DATABASES

In total, five databases were available for analysis. Three databases were issued from trials and two from cohorts. One database had semiflexed views, in this case an antero-posterior fluoroscopically assisted semiflexed view[15]. In all cases, the diagnosis of OA was based on the American College of Rheumatology criteria[21]. Characteristics of populations are shown Table I.

- The French database was obtained during a randomized placebo-controlled trial of laterally wedged insoles in 156 patients with symptomatic knee OA over a 24-month period[22]. The patients had been structurally evaluated using extended view knee X-rays.
- The Indianapolis database was obtained during a randomized placebo-controlled trial of doxycycline in obese women with unilateral knee OA over a 30-month period[23]. Each subject underwent radiographs at baseline and 30 months later, which included a fluoroscopically standardized semiflexed antero-posterior view of each knee[15] and a standing antero-posterior view. In the present study, the baseline semiflexed and extended views of 298 and 131 patients, respectively, were extracted and analyzed, as well as the 30-month follow-up semiflexed and extended views of 50 patients.
- The Houston database was obtained during a randomized placebo-controlled trial of acupuncture in 289 patients, structurally evaluated using extended views knee X-rays.
- The Toronto database was issued from a population-based cohort of symptomatic OA patients[24]. The patients had been structurally evaluated using extended view

Table I
Description of databases and participants' characteristics

Origin of data	Study design	Number of patients	Age, years, mean (±SD)	Gender (%female)	BMI, mean, Kg/m^2 (±SD)	WOMAC pain median (range)	WOMAC function median (range)	KL (% of patients)			OARSI (% of patients)			Cat. JSW (% of patients)		
								0–1	2	3–4	0–1	2	3	0–1	2	3
Paris, France	Trial, insoles vs placebo insoles	144	64.4 (11.7)	72.2	28.4 (5.1)	52 (15–95)	50 (7–87)	6	59	35	68	31	1	52	31	17
Indianapolis, US	Trial, doxycycline vs placebo	131 (extended views)	53.7 (5.4)	100	37.7 (6.5)	44 (20–100)	47 (20–100)	15	69	16	66	27	7	86	8	6
Indianapolis, US	Trial, doxycycline vs placebo	298 (semiflexed views)	54.3 (5.5)	100	36.7 (6.1)	40 (20–100)	42 (20–100)	5	66	29	52	31	17	54	31	15
Houston, US	Trial, acupuncture vs placebo	289	64.3 (8.9)	66.4	36.0 (10.9)	45 (5–100)	44 (0–100)	0	36	64	11	45	43		NA	
Johnston County, US	Population-based cohort	735	67.2 (9.5)	65.7	30.4 (6.5)	15 (0–100)	19 (0–100)	62	18	20	80	13	7	68	20	12
Toronto, Canada	Cohort	162	69.9 (8.4)	71.0	28.2 (5.3)	40 (0–95)	42 (0–84)	49	18	33	36	48	16		NA	

WOMAC: WOMAC scores are normalized 0–100; KL: KL radiographic grade; OARSI: OARSI joint space narrowing radiographic grade; NA: not available.

knee X-rays. In this cohort, 162 patients had complete clinical and radiographic data and were analyzed here.

Because the magnification factor for Houston and Toronto X-rays was not available, these X-rays were analyzed according to KL and OARSI scores but JSW could not be calculated.

- The Johnston County Osteoarthritis Project database is a community-based cohort with symptomatic and asymptomatic OA as well as non-OA participants[25]. The 735 participants structurally evaluated using extended view knee X-rays were analyzed.

CONSTRUCT VALIDITY

It should be noted that there were no available data for predictive validity, i.e., which radiographic score best predicts later total joint replacement. Construct validity (cross-sectional association with symptoms) is shown in Tables II and III. Association with symptoms was low, for all radiographic scores; only in the less advanced, less symptomatic Johnston County cohort was an association shown, and this association was similar for all three radiographic grading techniques. In addition, there was an association between pain and structure, assessed using categorical JSW, in the insole trial.

RELIABILITY

Reproducibility of readings is shown in Table IV. Inter-reader reliability was not obtained for semiflexed views. Inter-reader reliability was highest for categorical JSW. Intra-reader reliability was highest for categorical JSW on semiflexed views, and tended to be higher for extended vs semiflexed views.

SENSITIVITY TO CHANGE

Sensitivity to change is shown Table V. This was assessed in a group of patients with both extended and semiflexed views, from the Indianapolis database. Responsiveness was highest for categorical JSW on semiflexed views.

Discussion

The objective of the present study was to compare, using the OMERACT filter, different radiographic views and scoring techniques, to select a categorical outcome measure defining structural severity in knee OA, for use as part of a composite criterion in clinical trials.

This large study of X-ray grading in knee OA leads to two main conclusions. The first concerns scoring techniques. Although KL, OARSI stages and categorization of JSW all have similar construct validity, it appears that categorical JSW is more reproducible and more sensitive to change. The second conclusion concerns radiographic techniques. In this analysis, semiflexed views (in this case, antero-posterior fluoroscopically positioned semiflexed views) were found to be superior to extended views.

To our knowledge, this is the first study to compare in a large number of X-rays the different available categorical scoring techniques for all psychometric properties. Acknowledged methodology, as recommended by the OMERACT group of experts[18], was applied to compare the radiographic scoring techniques. However, it was not possible to evaluate all aspects of validity, in particular predictive validity. In addition, this study only concerned the tibio-femoral joint; we did not study the femoro-patellar joint. And finally only one type of semiflexed incidence was available for analysis.

In this study, there was no significant difference between the three radiographic scoring techniques with regards to construct validity, which was overall low. The observed discordance between clinical and radiographic data has been noted by several authors[26,27], while others published conflicting results[28,29]. Several hypotheses can be proposed for explanation: (1) X-rays may not effectively capture the pathologic features of OA that are responsible for pain and physical dysfunction, e.g., synovitis; (2) the current

Table II
Construct validity of the three knee X-ray grading systems: adjusted odds ratios to explain WOMAC total and WOMAC pain (in quartiles) by X-ray grade

Database	X-ray grade	WOMAC total			WOMAC pain		
		KL	OARSI	Cat. JSW	KL	OARSI	Cat. JSW
Paris	1 vs 0	NS	NS	NS	NS	NS	2.78 (0.86–8.98)*
	2 vs 0						3.77 (1.13–12.5)**
	3 vs 0						NS
Indianapolis (extended views)		NA	NA	NA	NS	NS	NS
Indianapolis (semiflexed views)		NA	NA	NA	NS	NS	NS
Houston		NS	NS	NA	NS	NS	NA
Johnston County	1 vs 0	NS	1.72 (1.2–2.45)***	NS	NS	1.74 (1.21–2.48)**	NS
	2 vs 0	1.74 (1.12–2.70)**	3.0 (1.78–5.07)○	1.53 (0.94–2.49)*	1.78 (1.14–2.77)**	3.68 (2.14–6.33)***	NS
	3 vs 0	2.45 (1.42–4.21)○	2.86 (1.38–5.94)***	2.82 (1.44–5.53)***	2.53 (1.46–4.39)**	4.39 (2.00–9.67)***	4.06 (1.97–8.4)***
	4 vs 0	2.7 (1.31–5.57)***			3.75 (1.73–8.14)***		
Toronto		NS	NS	NA	NS	NS	NA

Cat. JSW: categorical JSW (grade 0: ≥5 mm, grade 1: 3.5–4.9 mm, grade 2: 2–3.4 mm, grade 3: <2 mm); NA: not available. NS: not significant. *P value < 0.05; **P value < 0.01; ***P value < 0.001; ○P value 0.06–0.08.

Table III
Construct validity of the three knee X-ray grading systems: adjusted odds ratios to explain WOMAC function and WOMAC stiffness (in quartiles) by X-ray grades (see Table II for abbreviations)

Database	X-ray	WOMAC function			WOMAC stiffness		
		KL	OARSI	Cat. JSW	KL	OARSI	Cat. JSW
Paris		NS	NS	NS	NS	NS	NS
Indianapolis (extended views)		NS	NS	NS	NA	NA	NA
Indianapolis (semiflexed views)		NS	NS	NS	NA	NA	NA
Houston		NS	NS	NA	NS	NS	NA
Johnston County	1 vs 0	NS	1.82 (1.28–2.58)***	NS	NS	1.69 (1.18–2.43)**	NS
	2 vs 0	1.90 (1.23–2.92)**	3.03 (1.82–5.06)***	1.56 (0.96–2.54)○	1.74 (1.11–2.74)*	3.02 (1.75–5.21)***	NS
	3 vs 0	2.50 (1.47–4.25)***	3.27 (1.59–6.74)**	2.95 (1.51–5.78)***	2.07 (1.19–3.58)**	2.97 (1.38–6.38)**	2.48 (1.25–4.92)**
	4 vs 0	3.51 (1.69–7.28)***			3.02 (1.39–6.53)**		
Toronto		NS	NS	NA	NS	NS	NA

Cat. JSW: categorical JSW (grade 0: 5 mm, grade 1: 3.5–4.9 mm, grade 2: 2–3.4 mm, grade 3: <2 mm); NA: not available. NS: not significant. *P value < 0.05; **P value < 0.01; ***P value < 0.001; ○P value 0.06–0.08.

X-ray views and/or scoring systems may be inadequate; it has been suggested that the evaluation of associations between structural knee changes and pain and function needs to include an assessment of the patellofemoral joint compartment and individual radiographic features rather than a global severity score[30]; (3) current methods used to evaluate pain and/or function may be inadequate; (4) fluctuations of the disease severity over time (e.g., severe radiographic OA may not lead to severe symptoms at the moment of this cross-sectional assessment); (5) differences between studies in inclusion criteria. In the present study, we did not evidence a relationship between X-rays and symptoms, except in the Paris database in which WOMAC pain was associated with categorical JSW, and the Johnston County cohort, in which the relationship between all symptom measures and all structural measures was moderately strong. This community-based cohort included symptomatic and asymptomatic individuals with a wide range of symptomatic and structural variability. In addition, the radiographs from the Johnston County Osteoarthritis Project were read by a single, highly experienced musculoskeletal radiologist (JBR), with high intra-rater and inter-rater reliability[25]. This reduction in radiographic measurement error and the larger sample size of this cohort likely contributed to observed associations between symptoms and structure in this cohort. In the other databases, though the level of pain was not necessarily high, all patients had definite OA (Table I). Further studies are needed to better understand this issue.

It would have been interesting to compare criterion validity (prediction of future joint replacement) between the radiographic views and the scoring techniques. However, unfortunately, joint replacement rates were low in the Johnston County cohort and not available in the other databases. High KL grades have been associated with further joint replacement, both at the hip and at the knee level. To our knowledge no comparative study has been performed in the same patients to assess the radiographic scoring techniques comparatively as regards predictive validity.

Reproducibility was higher for categorical JSW than for the other scoring techniques, and responsiveness was higher for categorical JSW, in particular on semiflexed views. In a previous study, it was reported that, in contrary to KL, the reproducibility of joint space narrowing scales was influenced by the experience of the reader, particularly in the lateral compartment[31]. In another study, the intra- and inter-reader transversal and longitudinal reliability were found to be stronger for joint space measurement than for KL scale and for a joint space narrowing scale[32]. The joint space narrowing scale was, however, a 6-grade scale, slightly different from the 4-grade OARSI scale, and JSW measurement was analyzed as a continuous, rather than as a categorical variable, such as in the present study.

Table IV
Reliability. Reproducibility of the three knee X-ray grading systems

	KL		OARSI		Cat. JSW	
	Weighted kappa (95% CI)	ICC (95% CI)	Weighted kappa (95% CI)	ICC (95% CI)	Weighted kappa (95% CI)	ICC (95% CI)
Inter-reader (extended views)	0.56 (0.38–0.73)	0.72 (0.38–0.86)	0.48 (0.32–0.64)	0.66 (0.30–0.83)	0.86 (0.76–0.96)	0.92 (0.86–0.95)
Intra-reader (extended views)	0.61 (0.42–0.80)	0.72 (0.55–0.83)	0.71 (0.56–0.86)	0.82 (0.70–0.89)	0.83 (0.71–0.96)	0.89 (0.82–0.94)
Intra-reader (semiflexed views)	0.50 (0.25–0.75)	0.56 (0.33–0.72)	0.67 (0.53–0.82)	0.79 (0.65–0.87)	0.89 (0.80–0.99)	0.94 (0.89–0.96)

Reliability was assessed in the Indianapolis database. Inter-reader reliability was not assessed for semiflexed views. Intra-reader reliability was assessed on the same X-rays for extended and semiflexed views. See Table II for abbreviations.

Table V
Sensitivity to change assessed by SRM

	KL	OARSI	Cat. JSW
SRM (extended views)	0.10	0.15	0.2
SRM (semiflexed views)	0.22	0.34	0.49

See Table II for abbreviations. Sensitivity to change was assessed in the Indianapolis database.

These differences might also explain, as well as the absence of semiflexed views, that in this previous work, the responsiveness of the joint space narrowing scale was higher than the responsiveness of KL, and comparable to the joint space measurement responsiveness.

This study confirms that antero-posterior fluoroscopically positioned semiflexed views are preferable to extended views in knee OA, for the purposes of clinical trials, since these views allow a more sensitive change assessment of OA. Unfortunately, and this is a limit of this large study, there were no other semiflexed views available for analysis. Therefore, our conclusions only concern antero-posterior semiflexed views, as proposed by Mazzuca *et al.*[15]. Further studies are needed to assess semiflexed views comparatively. The studies published to date[6,11,17], comparing data from different flexion X-ray protocols, have not shown any clear advantage of a semiflexed technique over another, though head-to-head comparisons have yet to be performed.

Semiflexed views have sometimes been criticized because they are more complicated to obtain than extended views, i.e., there are feasibility issues, in particular regarding fluoroscopic positioning; however, once again, the main focus of this work was clinical trials, not clinical practice; semiflexed views are feasible in the context of clinical trials. Other elements to discuss regarding feasibility (as defined in the OMERACT filter) include training of persons to perform semiflexed views, and precise positioning in a reproducible angle of semiflexed knees; as for time and difficulty/training of scoring and cost, they are similar for extended and for semiflexed views.

Finally, it must be pointed out that other instruments evaluating the structural domain in OA are under evaluation, in particular MRI. In 2004, the working group considered that the available data were not sufficient to state that MRI could be used instead of plain X-rays. However, data on MRI are currently increasing, and this position might be reconsidered in the future.

In conclusion, this large study indicates that categorical JSW measurement might be the preferred instrument to evaluate structure in a set of criteria for use in clinical trials evaluating potential disease-modifying drugs in OA, and suggests that semiflexed views may present advantages over extended views. Further studies are required to evaluate predictive validity for total knee replacement, to assess whether a categorical or a dichotomized JSW measurement should be used in such set of criteria, and to establish final threshold(s).

Conflict of interest

The authors declare no conflict of interest.

The study sponsors played no role in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Acknowledgments

Sources of support: This work was supported by unrestricted grants from both scientific societies (OARSI and OMERACT) and pharmaceutical companies: Pfizer, Astra Zeneca, Novartis, and Negma Lerads laboratories. The Johnston County Osteoarthritis Project data collection was supported by the Centers for Disease Control and Prevention/Association of Schools of Public Health S043 and S3486, and the National Institute of Arthritis, Musculoskeletal and Skin Diseases Multipurpose Arthritis and Musculoskeletal Disease Center 5-P60-AR30701 and Multidisciplinary Clinical Research Center 5-P60-AR49465. The Houston database was supported by a grant from the National Institute for Arthritis, Musculoskeletal and Skin Diseases (NIAMS; R01 49999). The Canadian cohort data collection was supported by grants from the Canadian Institutes of Health Research (formerly, the Medical Research Council of Canada) (MT-12919).

References

1. Elders MJ. The increasing impact of arthritis on public health. J Rheumatol 2000;60(Suppl):6–8.
2. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol 1999;26:855–61.
3. Altman RD, Abadie E, Avouac B, Bouvenot G, Branco J, Bruyere O, *et al*. Group for Respect of Excellence and Ethics in Science (GREES). Total joint replacement of hip or knee as an outcome measure for structure modifying trials in osteoarthritis. Osteoarthritis Cartilage 2005;13:13–9.
4. Maillefert JF, Hawker GA, Gossec L, Mahomed NN, Lohmander S, Dieppe PA, *et al*. Concomitant therapy: an outcome variable for musculoskeletal disorders? Part 2: total joint replacement in osteoarthritis trials. J Rheumatol 2005;32:2449–51.
5. Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R, *et al*. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol 2007;34:1432–5.
6. Mazzuca SA, Brandt KD, Buckwalter KA. Detection of radiographic joint space narrowing in subjects with knee osteoarthritis: longitudinal comparison of the metatarsophalangeal and semiflexed anteroposterior views. Arthritis Rheum 2003;48:385–90.
7. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.
8. Altman RD, Hochberg M, Murphy WA, Wolfe F, Lequesne M. Radiographic atlas for osteoarthritis of the hand, hip and knee. Osteoarthritis Cartilage 1995;3(Suppl A):3–19.
9. Brandt KD, Mazzuca SA, Conrozier T, Dacre JE, Peterfy CG, Provvedini D, *et al*. Which is the best radiographic protocol for a clinical trial of a structure modifying drug in patients with knee osteoarthritis? J Rheumatol 2002;29:1308–20.
10. Altman RD, Fries JF, Bloch DA, Carstens J, Cooke TD, Genant H, *et al*. Radiographic assessment of progression in osteoarthritis. Arthritis Rheum 1987;30:1214–25.
11. Piperno M, Hellio Le Graverand MP, Conrozier T, Bochu M, Mathieu P, *et al*. Quantitative evaluation of joint space width in femorotibial osteoarthritis: comparison of three radiographic views. Osteoarthritis Cartilage 1998;6:252–9.
12. Mazzuca SA, Brandt KD, Buckland-Wright JC, Buckwalter KA, Katz BP, Lynch JA, *et al*. Field test of the reproducibility of automated measurements of medial tibiofemoral joint space width derived from standardized knee radiographs. J Rheumatol 1999;26:1359–65.
13. Duddy J, Kirwan JR, Szebenyi B, Clarke S, Granell R, Volkov S. A comparison of the semiflexed (MTP) view with the standing extended view (SEV) in the radiographic assessment of knee osteoarthritis in a busy routine X-ray department. Rheumatology (Oxford) 2005;44:349–51 (Epub 2004 Nov 16).
14. Vignon E, Piperno M, Le Graverand MP, Mazzuca SA, Brandt KD, Mathieu P, *et al*. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon schuss views. Arthritis Rheum 2003;48:378–84.
15. Mazzuca SA, Brandt KD, Buckwalter KA, Lequesne M. Pitfalls in the accurate measurement of joint space narrowing in semiflexed, anteroposterior radiographic imaging of the knee. Arthritis Rheum 2004; 50:2508–15.
16. Conrozier T, Mathieu P, Piperno M, Favret H, Colson F, Vignon M, *et al*. Selection of knee radiographs for trials of structure-modifying drugs in

patients with knee osteoarthritis: a prospective, longitudinal study of Lyon schuss knee radiographs with the definition of adequate alignment of the medial tibial plateau. Arthritis Rheum 2005;52:1411–7.
17. Radiography Working Group of the OARSI–OMERACT Imaging Workshop, Le Graverand MP, Mazzuca S, Lassere M, Guermazi A, Pickering E, *et al*. Assessment of the radioanatomic positioning of the osteoarthritic knee in serial radiographs: comparison of three acquisition techniques. Osteoarthritis Cartilage 2006;14(Suppl A):A37–43.
18. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–9.
19. Bellamy N. WOMAC Osteoarthritis Index. A User's Guide. Ontario: London Health Services Centre, McMaster University; 1996.
20. Kendall MC, Stuart A. The Advanced Theory of Statistics. New York: Hafner Publishing 1967;Volume II. pp. 302–4.
21. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
22. Maillefert JF, Hudry C, Baron G, Kieffert P, Bourgeois P, Lechevalier D, *et al*. Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis: a prospective randomized controlled study. Osteoarthritis Cartilage 2001;9:738–45.
23. Brandt KD, Mazzuca SA, Katz BP, Lane KA, Buckwalter KA, Yocum DE, *et al*. Effects of doxycycline on progression of osteoarthritis: results of a randomized, placebo-controlled, double-blind trial. Arthritis Rheum 2005;52:2015–25.
24. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, *et al*. Differences between men and women in the rate of use of hip and knee arthroplasty. N Engl J Med 2000;342:1016–22.
25. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, *et al*. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol 2007;34:172–80.
26. Cicuttini FM, Baker J, Hart DJ, Spector TD. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage 1996;4:143–7.
27. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000;27:1513–7.
28. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum 1987;30:914–8.
29. Lethbridge-Cejku M, Scott WW Jr, Reichle R, Ettinger WH, Zonderman A, Costa P, *et al*. Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. Arthritis Care Res 1995;8:182–8.
30. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, *et al*. Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum 2006;54:230–5.
31. Gunther KP, Sun Y. Reliability of radiographic assessment in hip and knee osteoarthritis. Osteoarthritis Cartilage 1999;7:239–46.
32. Ravaud P, Giraudeau B, Auleley GR, Chastang C, Poiraudeau S, Ayral X, *et al*. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol 1996;23:1756–64.